

THE GENOMIC DIVERSITY AND PHENOTYPE CONNECTION



Terry M. Casstevens¹ and Edward S. Buckler²

¹Departments of Statistics and Genetics, North Carolina State University, Raleigh, NC 27695-7566,
²USDA-ARS, Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614

<http://www.maizegenetics.net>

1. Introduction

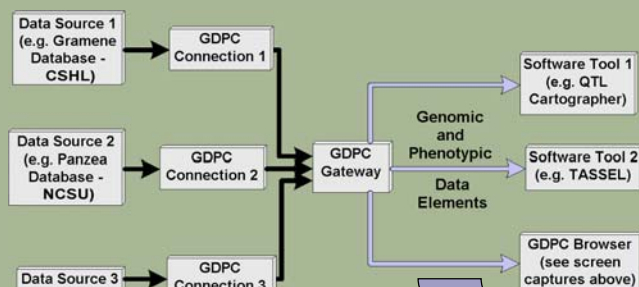
The motivation for the Genomic Diversity and Phenotype Connection, GDPC, stems from the increasingly large genomic and phenotypic datasets that are becoming available in plant biology. Databases and software tools that integrate information on genomic diversity with phenotypic data provide an opportunity to help bridge genomics and plant breeding. GDPC provides a JAVA interface that facilitates access to that data and has the following advantages:

- Simplifies access to data.
- New data sources are easily made available.
- Analysis tools require no additional software development to use new data sources.
- Analysis tools can access multiple data sources simultaneously.

Early version at: <http://www.maizegenetics.net>.

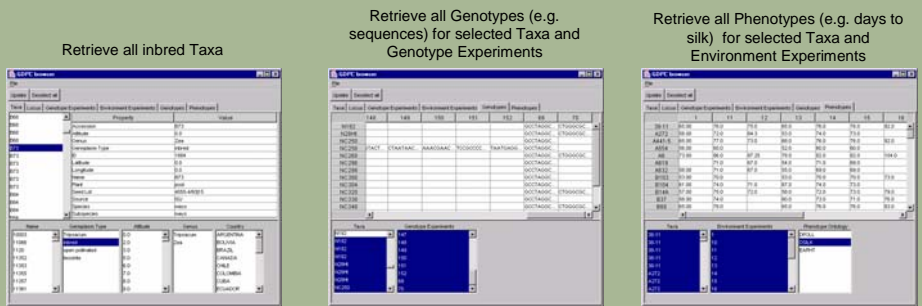
2. Data Flow

This diagram shows how software tools can use multiple data sources via the gateway provided by GDPC. All information is requested from the GDPC gateway (JAVA interface). The GDPC gateway retrieves the requested data from the data sources via special connections designed specifically for each supported data source. The gateway then forwards the "Genomic and Phenotypic Data Elements" from all its managed data sources to the requesting software tool. These software tools can operate with any existing or future connection without any further coding.



4. GDPC Browser

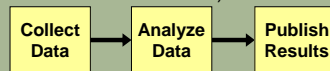
The GDPC Browser is a software tool which uses the GDPC Gateway. It retrieves Genomic and Phenotypic data based on selected criteria.



5. Genomic Diversity and Phenotypic Data Life Cycle

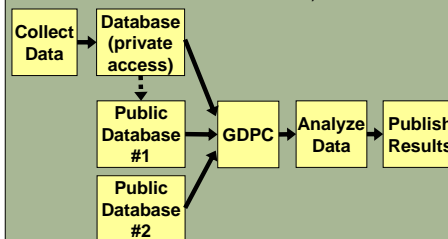
Typical Procedure

(Original data typically unavailable to other researchers)



Proposed Procedure

(Data can be made available for reanalysis by other researchers)



3. Genomic and Phenotypic Data Elements

- **Environment Experiment:** This defines when and where the phenotypic evaluation was carried out.
- **Environment Experiment Group:** For example, all environment experiments harvested during a particular range of dates.
- **Genotype Experiment:** This defines how the sequence, SNP, SSR, etc. data was collected.
- **Genotype Experiment Group:** For example, all genotype experiments performed on a particular locus.
- **Genotype Table:** Table of genotype data (genetic characteristics of an organism) for a given list of genotype experiments and list of taxa.
- **Locality:** Description of a particular geographic location.
- **Locus:** A specific location on a chromosome. (e.g. gene, etc.)
- **Locus Group:** For example, all loci located on chromosome 3.
- **Phenotype Ontology:** Defines the physical trait of an organism.
- **Phenotype Table:** Table of phenotype data for a given phenotype ontology, list of environment experiments, and list of taxa.
- **Taxon:** Defines the source of the germplasm used in the experiment.
- **Taxon Group:** For example, all inbred taxa.

The diagram below illustrates the various genomic and phenotypic data elements accessible through the GDPC Gateway. The lines connecting the data elements show their relationship, such that the element next to the diamond contains the element pointed to by the arrow. The notation, "1..n", represents a "one to many" relationship, and its absence indicates a "one to one" relationship. For example, one or more taxa are components of a taxon group.

